Improving the Performance of Communication-Intensive Parallel Applications Executing on Clusters

Xiao Qin and Hong Jiang

Department of Computer Science and Engineering
University of Nebraska-Lincoln

Clusters have emerged as a primary and cost-effective infrastructure for parallel applications, including communication-intensive applications that transfer a large amount of data among nodes of a cluster via the interconnection network. Conventional load balancers have been proven effective in increasing the utilization of CPU, memory, and disk I/O resources in a cluster. However, most of the existing load balancing schemes ignore network resources, leaving open the opportunity for significant performance bottleneck to form for communication-intensive parallel applications due to unevenly distributed communication load. To remedy this problem, we propose a communication-aware load balancing technique that is capable of improving the performance of communication-intensive applications by increasing the effective utilization of network resources in clusters. To facilitate the proposed load-balancing scheme, we introduce a behavior model for parallel applications with large requirements of CPU, memory, network, and disk I/O resources. The proposed load-balancing scheme can make full use of this model to quickly and accurately determine the load induced by a variety of parallel applications. Simulation results on executing a diverse set of both synthetic bulk synchronous and real parallel applications on a cluster show that the proposed scheme can significantly improve the performance both in slowdown and turn-around time over three existing schemes by up to 206% (with an average of 74%) and 235% (with an average of 82%), respectively.

0-7803-8694-9/04/$20.00 ©2004 IEEE

Authorized licensed use limited to: Auburn University. Downloaded on November 18,2026 at 18:27:18 UTC from IEEE Xplore. Restrictions apply.